Catégorisation par objectifs de la visualisation pour la sécurité

Christopher Humphries, Nicolas Prigent, Christophe Bidan, Frédéric Majorczyk

Catégorisation par objectifs de la visualization pour la sécurité

C. Humphries, N. Prigent, C. Bidan, and F. Majorczyk

[1] Inria, Équipe Cidre prenom.nom@inria.fr
[2] Supélec, Équipe Cidre prenom.nom@supelec.fr
[3] Supélec, Équipe Cidre prenom.nom@supelec.fr
[4] DGA-MI prenom.nom@intradef.gouv.fr

La visualisation est désormais une fonctionnalité fréquente dans les outils de sécurité (Fig.1) qui a été appliquée sur de nombreux types de données : événements réseaux, données système, analyse statique de binaires, analyse de la structure de *malware*, par exemple.

Dans le cadre de cet article, nous nous intéressons à la visualisation dans le domaine de la securite, et plus spécifiquement, la securite des reseaux. Nous nous concentrons donc sur les evenements reseau, et non pas sur l'analyse statique de binaires, ni sur la structure de malwares.

Dans son état actuel, la visualisation pour la sécurité des systèmes d'information est plus souvent le résultat de l'application expérimentale de techniques de visualisation venant d'autres domaines sur des données de sécurité montrant des problèmes similaires. Par exemple, une visualisation capable de bien afficher une hiérarchie sera utilisée pour les systèmes de fichiers et les adresses ; les coordonnées parallèles [5] et les nuages de points seront utilisés pour faire de la corrélation d'événements réseau ; les *sparklines* afficheront efficacement des métriques et les graphes de nœuds seront souvent utilisés pour les réseaux, à la fois physiques et sociaux.

Les exemples les plus marquants en visualisation pour la sécurité inspirés par d'autres domaines viennent des outils d'analyse biologique. Ainsi, Circos, utilisé initialement pour l'analyse de données génomiques, a été adapté pour la conscience de la situation [LAMF05,FA07] et l'analyse de communications par email. Nous pouvons également citer les *hive plots* [hiv13], utilisés en tant qu'alternative au graphe de noeuds pour représenter les transferts de protéines dans les bactéries, et qui ont été ré-appliquées pour la visualisation de calculs en mémoire distribuée [EW12].

En étudiant les différents outils de visualisation pour la sécurité, nous avons identifié trois catégories dépendant de l'objectif visé. En premier lieu, les outils de visualisation pour la **supervision** des serveurs ou des réseaux ont pour objectif de surveiller certaines métriques du système d'information en vue de détecter au plus tôt des anomalies. En second lieu, les outils de **fouille visuelle** permettent

5. Un graphe par coordonnées parallèles affiche un axe par champ de données, puis lie les variables par rapport à ces axes.

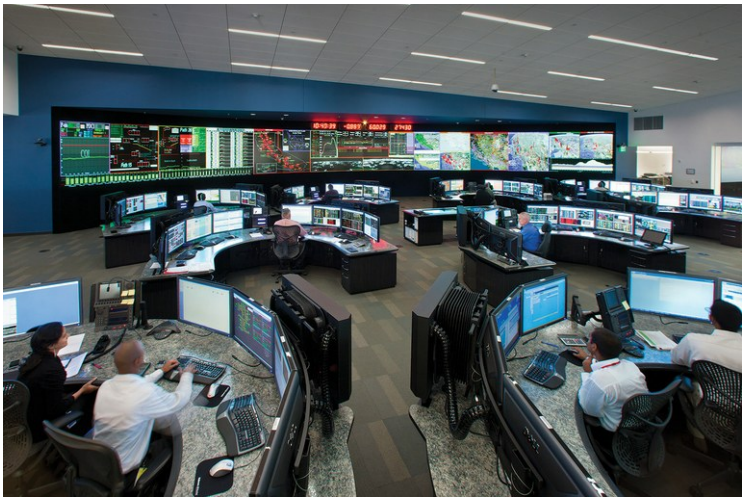

Figure 1. Le centre de controle du “California Independent System Operator", qui gère 80% de l'énergie de l'état.

d'explorer et d'analyser les données de sécurité pour expliquer les anomalies et identifier les scénarios d'attaque, ou encore localiser des intrusions qui auraient été manquées. Enfin, les outils de visualisation utilisés pour établir un **rapport** facilitent la compréhension des événements et de leurs implications. Ces trois catégories ne sont bien évidemment pas disjointes, certains outils de visualisation pouvant avoir plusieurs objectifs.

Cet article présente un etat de l'art des outils de visualisation pour la securite des systemes d'information et une classification des outils basee sur les objectifs. La section 1 présente les outils utilisés pour la surveillance de réseaux ou de systèmes. La section 2 présente les outils relatifs à la fouille visuelle de données liées à la sécurité. Les outils les moins nombreux sont ceux dédiés à la rédaction de rapport visuel ; ils sont présentés dans la section 3. Finalement, nous discutons notre choix de classification et concluons.

1 Visualisation pour la surveillance

La surveillance de systèmes est une des utilisations de la visualisation la plus commune en sécurité. Il s'agit de s'assurer que les systèmes fonctionnent de façon nominale. Dans ce but, la visualisation est utilisée pour détecter des changements alarmants ou des signes clairs d'intrusion. La bonne conception de l'outil de visualisation, notamment dans le choix des données et de leur représentation, permet de capturer un groupe spécifique de motifs. Les tableaux de bords sont des représentations adaptées pour observer des tendances ou repérer des valeurs

spécifiques. Cependant, des configurations visuelles plus complexes sont nécessaires pour capturer des motifs plus evasifs et des corrélations suspectes. De ce fait, les outils de surveillance sont généralement conçus pour résoudre un problème en particulier. Ils sont bien adaptés à la réalisation de cette tâche, mais ne sont généralement pas flexibles. Nous classons les outils de visualisation pour la supervision en trois catégories distinctes suivant leur objectif : la corrélation rapide, la décision rapide et le passage à l'échelle.

1.1 Corrélation rapide

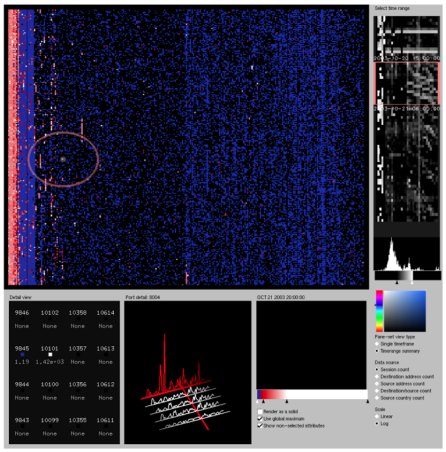

Figure 2. PortVis [MKL04], un example de visualisation en nuage de points.

Les outils de corrélation rapide permettent la recherche de motifs et de signes d'anomalies en faisant des corrélations ou en reconnaissant des motifs visuels.

Snortview [KO04] permet une corrélation visuelle simple pour gérer les faux positifs émis par une sonde de détection d'intrusions Snort [Sno13]. Les alertes sont tracées suivant leur date et leur adresse source ; leur type et leur priorité sont également affichés. Cette représentation a pour objectif de corréler la fréquence et le nombre d'alertes pour déterminer les vrais positifs dans le flot d'alertes émises par la sonde. Ceux-ci se démarquent des tendances et deviennent ainsi plus visibles, ce qui permet de se concentrer sur des nouveaux motifs d'attaques.

Colombe et al. proposent une méthode de définition de profils visuels statistiques [CS04] similaires en apparence aux visualisations à vue « galaxie », et dans lesquels les alertes sont des rangées entières de pixels, codés par couleur selon le temps. Chaque colonne est construite suivant des étiquettes associées aux alertes. Cette méthode permet de comparer les propriétés des alertes et de détecter des alertes inhabituelles ou inconnues.

PortVis [MKL04] est un autre outil qui utilise des visualisations en nuage de points. Différents événements de sécurité sont représentés par des pixels suivant le temps. Deux vues globales sont utilisées : la première utilise un axe de temps général (Fig. 2), l'autre un axe de temps avec une échelle horaire. Ensemble, ces vues permettent une détection d'événements périodiques. Des vues détaillées plus petites affichent les ports concernés et un graphe de leur historique avec

des paramètres de gradients de couleurs. La combinaison de ces multiples vues globales et détaillées de façon synchronisée permet à l'outil d'agir comme un filtre multi-modal ainsi qu'une exploration en profondeur.

Utilisant une visualisation en carte de points, IDSRainstorm [AL05,AC06] a été conçu pour afficher un grand nombre d'alertes (une journée entière d'alertes sur le réseau de GeorgiaTech). Une première visualisation représente par des points les alertes ; la couleur du point permet de déterminer la sévérité. L'axe vertical permet de déterminer à quel groupe d'adresses sont reliées les alertes alors que l'axe horizontal est un axe de temps représentant une journée. Un pixel agrège plusieurs alertes et adresses et représente l'alerte la plus sévère dans ce groupe pour cette période de temps. Pour inspecter une zone d'intérêt particulière, une seconde visualisation fournit une vue agrandie de la zone sélectionnée. Ce mode plus détaillé affiche chaque alerte sans les grouper par adresse, et affiche les connexions à partir d'adresses externes pour indiquer les alertes déclenchées par des hôtes externes.

IPMatrix [Koi05] propose également une visualisation basée sur deux nuages de points en coordination : le premier nuage pour les espaces A et B d'adressage au « niveau internet » et le second pour les espaces C et D au « niveau local ». Chaque nuage affiche les attaques détectées selon l'espace d'adressage en tant que pixels colorés suivant un code correspondant au type d'attaque. Ces points sont tracés sur une grille de cases d'agrégation affichant le nombre d'attaques par bloc d'adresses. Des histogrammes accompagnent chaque vue pour assister et ameliorer la lisibilite des chiffres des attaques. Une version tridimensionnelle de l'outil est également disponible : elle empile les representations des espaces et utilise des cartes de hauteur plutot que les cases d'agregation pour afficher les densites des attaques.

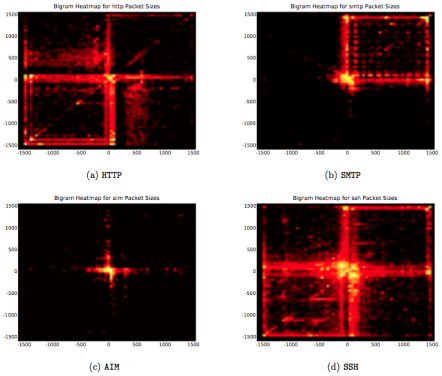

Figure 3. Quatre protocoles représentés avec des motifs visuels formés par des visualisations par nuages de points.

Avec l'objectif de détecter les scans, Irwin et al. utilisent aussi des nuages de points en 3D [IVR]. Leur outil affiche les connexions d'une adresses source à une adresse et un port de destination. Les connexions disparaissent avec le temps, ce qui permet de voir lesquelles sont les plus récentes. En utilisant cette combinaison de techniques, des millions de connexions peuvent être représentées en quasi temps-réel, ce qui permet de discerner visuellement les motifs de scan.

Dans le but de créer des visualisations reconnaissables caractérisant les protocoles réseau, Wright et al. construisent des motifs visuels [WMM03] en utilisant des nuages de points. De manière similaire aux cartes de chaleur, ils tracent la taille des packets sur un axe temporel (Fig. 3), et réussissent à montrer que des protocoles tels que SSH et HTTP sont caractérisés par des motifs très différents avec cette technique. De la même manière, se basant sur l'hypothèse que les protocoles au niveau internet ont des comportements temporels et des tailles de packets significatifs, Lian et al. utilisent des cartes de chaleur pour eux aussi fabriquer des motifs visuels [LMM10] pour différents protocoles. Chaque représentation trace un chemin basé sur la taille du paquet et le signe de sa direction : positif pour client vers serveur, négatif dans l'autre direction.

Contrairement aux outils présentés précédemment, VisFlowConnect [YA04] utilise des coordonnées parallèles pour faire apparaitre des corrélations en combinant la visualisation avec un curseur pour le temps et des filtres pour l'exploration de flux réseau. Cette visualisation facilite la corrélation multidimensionnelle de flux réseaux. Il faut noter cependant que la découverte de motifs intéressants dépend fortement de l'ordre d'arrangement des dimensions dans la représentation.

Visual Firewall [LC05] utilise une combinaison de plusieurs fenêtres de visualisation pour la détection et la reconnaissance de motifs en relation avec les pare-feu dans le but d'aider à leur configuration. Une vue en temps-réel des échanges permet d'afficher les paquets allant d'adresses externes à des ports locaux. Les paquets qui traversent le pare-feu génèrent un code de statut, alors que les paquets rejetés rebondissent. Une visualisation a coordonnees paralleles affiche l'historique de ces échanges et permet leur correlation avec le temps. Une seconde visualisation permet d'affiche des graphes de mesures sur les flux entrants et sortants. Enfin, une derniere visualisation a coordonnees paralleles permet de relier les connexions entre des sous-reseaux distants avec des types d'alertes, un axe vertical permettant d'indiquer la date de l'alerte et sa severite à l'aide de la coloration du point.

Muelder et al. [MMB05] proposent une méthodologie de visualisation ayant deux vues synchonisées pour déterminer des motifs dans le trafic réseau. Un graphe global affiche les relations entre les nœuds du réseau. La visualisation secondaire plus détaillée propose une représentation par traces de scan par rapport à deux espaces d'adressage, colorés suivant la date du scan, et combinés avec des « scalograms » (c.-à-d. des histogrammes mis à l'échelle) pour faciliter la comparaison de motifs. Cette configuration fournit un cycle rapide entre la vue globale et les vues détaillées pour comparer différentes motifs réseau typiques.

Comme nous le montrent les outils présentés précédemment, les nuages de points et les coordonnées parallèles ont été très utilisés pour permettre une corrélation rapide entre des événements. Quand des motifs et des signes d'anomalies sont détectés, l'inspection rapide des données est facilitée pour permettre une compréhension rapide de la situation.

1.2 Décision rapide

Avec l'objectif d'améliorer la connaissance de la situation, Livnat et al. présentent VisAlert [LAM+05,LAMF05,FA07], un système de visualisation radiale innovant qui vise à répondre rapidement à trois questions : que s'est-il passé, quand et où ? Des alertes sont tracées sur une tranche colorée radiale selon le type et se déplacent vers l'extérieur avec l'âge. Pour localiser l'alerte, l'espace à l'intérieur des anneaux héberge une visualisation spatiale ou organisationnelle à laquelle chaque alerte est liée. Par exemple, quand on visualise des intrusions réseau, cette visualisation centrale peut être une représentation en graphe de nœuds du réseau. On peut noter que cet outil a été utilisé dans d'autres domaine tels que la gestion des désastres et les appels aux urgences.

Pour fournir un point de départ aux analystes réseau, Overflow [GBT+09] propose une composition de trois visualisations. La première est une visualisation radiale affichant les différents éléments du réseau. Des lignes montrent les communications entrantes et sortantes entre. La seconde visualisation affiche la hiérarchie réseau détaillée pour l'élément sélectionné en utilisant un *tree map*, avec le même code couleur que l'affichage simplifié en anneaux. La dernière visualisation montre les groupes d'adresses IP pour chaque élément réseau.

Pour améliorer la réactivité et multiplier les options disponibles pour les opérateurs de sécurité, Hertzog propose une nouvelle stratégie de construction de visualisations [Her06]. La première étape est de réduire le nombre de données aux plus importantes pour réduire la charge de stockage, et ensuite de regrouper les données. Par exemple, les applications utilisées pour naviguer sur l'Internet peuvent être agrégées. Pour illustrer cette stratégie, deux visualisations sont présentées. La première est une visualisation interactive en coordonnées parallèles affichant les connexions d'un utilisateur par source, application, port et destination. Plusieurs nœuds sont regroupés pour simplifier le graphe, et des chemins spécifiques peuvent être isolés de façon interactive. La deuxième visualisation utilise un tracé bidimensionnel de l'utilisation de l'application selon le temps. Des segments d'utilisation sont colorés quand ils correspondent à des alertes. Des histogrammes affichent le niveau de ces alertes et fournissent des détails à la demande pour chaque segment.

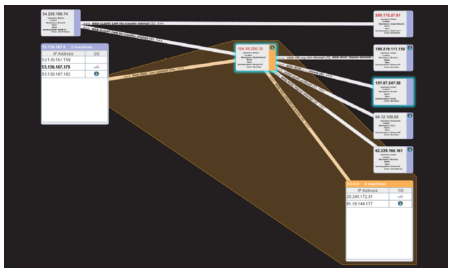

Figure 4. Représentation par cartes d'hôtes dans NIMBLE [RER+10].

Les administrateurs de systèmes ont parfois besoin de prendre des décisions rapidement. Rasmussen et al. proposent l'outil NIMBLE [RER+10] qui repose sur une visualisation en graphe de nœuds qui fait des recommandations basées

sur un apprentissage automatique. Des cartes sont affichées et reliées suivant le déclenchement d'alertes. Chaque carte représente un ou plusieurs hôtes, avec plus ou moins de détails, et les liens apportent ensuite une description sur le type de l'alerte concernant les deux noeuds (Fig. 4). Les explications sont affichées dans une liste à côté, et lors de la sélection d'un des éléments de la liste, les nœuds et les connexions concernés sont mis en évidence.

Pour permettre d'améliorer encore la connaissance de la situation, les outils de visualisation permettant la décision rapide dans la surveillance intègre souvent de nombreuses sources de données. Ils peuvent ainsi représenter de nombreux aspects de systèmes entiers d'hôtes et de processus. Ils utilisent souvent des graphes de nœuds et diminuent ainsi la distance mentale référent-référé. Ces solutions ont besoin de maintenir un état fonctionnel pour gérer des quantités de données grandissantes en temps-réel.

1.3 Passage à l'échelle

La visualisation en temps-réel est d'habitude seulement utilisée pour des ensembles de données filtrées. Pour améliorer la visualisation de données réseau en temps-réel, Daniel et al. proposent deux visualisations [DBWW10]. Le premier, nommé CLIQUE, est basé sur le projet LiveRac [6] et fournit une représentation tabulaire de données avec des colonnes pour chaque service et des rangées pour les utilisateurs. Chaque cellule contient initialement une visualisation de type *sparkline* qui peut être agrandie pour afficher plus de détails. La deuxième visualisation est un tracé radial similaire à un affichage radar conçu pour les grands affichages à haute résolution qui représente les flux réseau pour une période spécifique de temps. Chaque flux est représenté par un pixel coloré placé à un angle correspondant à sa date et avec une position radiale paramétrable.

Kintzel et al. [KFM11] proposent quatre visualisations pour surveiller des grands nombres d'hôtes, dans un ordre qui permet un filtrage et une focalisation progressifs. La première représentation basique utilise un graphe de type *sparkline* ou en barres affichant l'activité de chaque hôte sur 24 heures. La seconde, appelée ClockView, représente chaque hôte par un graphe radial, similaire à une horloge, affichant de nouveau l'activité sur 24 heures, mais en tant que *glyphe* permettant ainsi de comparer les différents hôtes. Pour la perception de la structure réseau, un graphe des communications peut être superposé à cette vue. Il est possible de se focaliser sur certains hôtes : la vue globale est alors remplacée par d'autres visualisations plus adaptées. L'activité est affichée en utilisant une matrice de pixels avec une granularité plus fine sur le temps, et les graphes « horloge » sont affichés en cordonnées parallèles pour permettre des corrélations. Une vue encore plus détaillée affiche une matrice de ports pour examiner les interactions entre deux machines spécifiques.

6. LiveRac est un outil de visualisation utilisé pour explorer des données à dimensions nombreuses pour un grand nombre d'hôtes.

Se basant sur la représentation en graphe de nœuds, Pearlman et al. utilisent des *glyphes* composés [PR] pour représenter les services en fonctionnement. Ici chaque noeud représente les services qu'il héberge avec un diagramme radial hierarchique. Ces services sont affichés de façon proportionelle, avec des anneaux externes représentant l'état le plus récent, ce qui permet de voir un historique à court-terme. Les hôtes simples ont des représentations simples, et peuvent être représentés en plus petit, ce qui permet de mettre des hôtes plus importants en évidence.

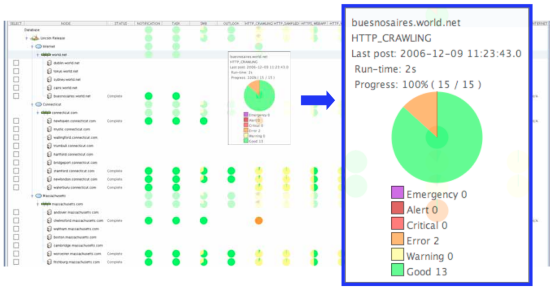

Figure 5. Utilisation de grilles de diagrammes circulaires pour visualiser des bancs d'essai de services [YFB+07].

Dans l'objectif de gérer et de surveiller des grands bancs d'essai d'hôtes qui hébergent de multiples services, Yu et al. proposent un outil de visualisation de bancs d'essai [YFB+] avec des arbres pour afficher la hiérarchie, des vues en graphe de nœuds et en matrices pour afficher les flux réseau, des chronologies pour afficher l'activité utilisateur, mais aussi une visualisation d'état d'appareils qui utilise des grilles de diagrammes circulaires multiples. Chaque diagramme représente l'état actuel du service et les proportions d'alertes, et fournit des détails sur demande.

Les outils de surveillance font une utilisation significative de visualisations permettant de percevoir des motifs (nuages de points, coordonnées parallèles), synchronisées et adaptées à la parallélisation et qui évoluent souvent en temps-réel.

2 Visualisation pour l'analyse

L'analyse de données est une étape nécessaire quand les outils de surveillance ont échoué ou quand aucune explication n'est disponible pour une anomalie. Les outils de visualisation adaptés à l'analyse permettent aux opérateurs de mieux comprendre les situations et les processus qui y ont mené. Analyser des données implique l'exploration de plusieurs configurations et plusieurs types de visualisations pour voir lesquels afficheront des résultats. L'analyse de données est générale guidée par un objectif spécifique tel que la recherche de tentatives d'intrusions répétées sur un système donné, ou encore la recherche de motifs ou signes d'activités potentiellement malveillantes

Les outils de visualisation adaptés à l'analyse utilisent des cycles de recherche similaires à ceux utilisés pour la surveillance avec cependant plus de contrôles dans la profondeur de recherche.

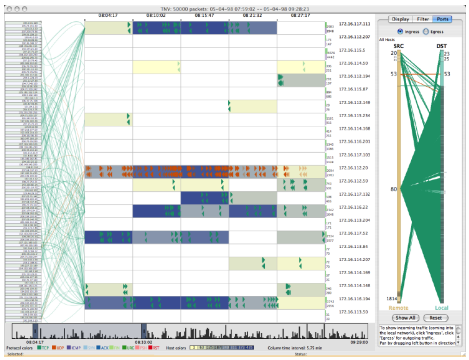

Figure 6. The Network Visualiser (TNV) [GL05] vise à offrir une image plus générale de captures de paquets réseau.

Des outils tels que « The Analyst's Notebook » [Ana] fournissent des visualisations adaptées à l'analyse, en utilisant des cartes pour localiser des données et un graphe de nœuds pour visualiser les liens entre ces données. En utilisant ces outils, un opérateur adoptera un cycle de raffinement entre chaque outil, en utilisant chaque outil comme un filtre pour les autres jusqu'à ce que l'information essentielle soit trouvée.

NVisionIP [LAL04] offre plusieurs vues pour explorer des données NetFlow. Au niveau de détails le plus large, une vue « galaxie » affiche les flux par sous-réseau et par hôte ; ces flux sont colorés suivant des groupes paramétrables. Quand une activité suspecte est visible, l'opérateur peut zoomer sur plusieurs vues plus petites qui comparent différents hôtes, puis sur une vue machine plus spécifique. Cette approche en profondeur est typique d'un outil de visualisation pour l'analyse. Deux articles complémentaires décrivent des extensions pour l'outil. Le premier [LSYN05] présente la capacité à enregistrer des chemins d'exploration visuelle et créer des règles correspondant aux motifs quand de nouvelles attaques sont découvertes. Le deuxième [Yur06] décrit la possibilité de comparer différents fichiers logs avec une vue des différences, des graphes linéaires pour représenter de densité de données, et des tracés de formes cherchent à mettre à profit les principes de perception *Gestalt*[7] afin de faciliter la reconnaissance de motifs dans les fichiers logs.

Au lieu de proposer des visualisations séquentielles pour explorer des données d'attaques Sybil dans des réseaux WiFi, Harrison et al. [HLW10] proposent un outil utilisant de multiples vues coordonnées. Chaque vue change de façon synchrone par rapport aux autres et agit comme un filtre pour l'interaction utilisateur. Pour l'analyse spatiale, la première vue propose un arrangement en graphe de nœuds du réseau. Pour l'analyse temporelle, un histogramme temporel affiche tous les événements sur une période donnée. La dernière vue est un nuage de points, pour visualiser des dimensions configurables à partir d'une analyse spectrale de données. En utilisant ces trois vues, l'utilisateur suit un processus de filtrage incrémental et trouve graduellement des points d'intérêt.

BGP Eye [Ran03] vise à aider la détection d'anomalies au niveau de BGP (Border Gateway Protocol) en utilisant quatre visualisations. La première visualisation est un graphe de nœuds qui affiche des événements déclenchés par des systèmes

7. La psychologie Gestallt décrit notre capacité à percevoir des liens et des groupes de formes dans une image avant de se concentrer sur celles-ci individuellement.

autonomes (AS) spécifiques. La deuxième utilise un arrangement spatial basé sur la distance du chemin avec les nœuds sources en bas, des nœuds puits en haut et tous les noeuds intermédiaires placés suivant leur distance. La troisième visualisation utilise un diagramme circulaire pour afficher les connexions entre les routeurs d'observation dans un anneau interne et les routeurs paires dans un anneau externe. La dernière vue est une association complexe de deux visualisations planaires : la première affiche les routeurs concernés, la seconde les statuts préfixés correspondants. Ceci permet de voir quels routeurs ont récemment vus des changements de chemins de routage.

D'autres outils de visualisation pour l'analyse ont adopté une approche différente de celle des cycles de recherche. Leur approche est basée sur une narration simple qui permet de montrer la progression des événements de façon détaillée.

TNV [GL05] vise à offrir une adaptation visuelle de l'outil d'inspection de paquets Wireshark pour obtenir une image plus générale de captures de paquets réseau. Une matrice centrale de visualisation affiche les paquets, annotés par direction et colorés par type de protocol. De chaque côté, les adresses sont listées et liées aux différents flux de paquets. À droite, une visualisation par coordonnées parallèles affiche les ports utilisés, et en dessous, un histogramme affiche l'évolution du trafic réseau qui permet le filtrage. Les filtres suivent une progression horizontale des paquets, ce qui rend la compréhension du trafic réseau plus facile.

Avec une approche similaire de circulation de gauche à droite, Portall [FMN05] permet à l'utilisateur d'explorer les processus client et serveur et la manière dont ils communiquent entre eux. Les clients sont listés sur la gauche, les serveurs sur la droite, et les liens entre processus sont représentées par un graphe de nœuds, affichés en tant que boites avec les détails du processus, notamment un histogramme d'activité de connexion.

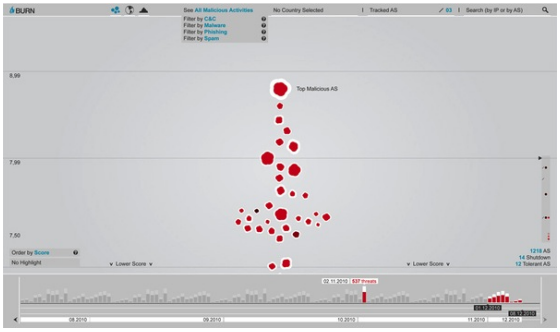

Figure 7.
BURN [RCDM+11] affiche et ordonne des systèmes autonomes par activité malveillante, en faisant une utilisation poussée d'animation et de transitions.

Roveta et al. propose BURN (*Baring Unknown Rogue Networks*) [RCDM+11] qui prend en entrée des alertes concernant des systèmes autonomes (AS). Cet outil utilise des techniques d'animation et des effets de transparence afin de faciliter

l'identification des comportements irréguliers (Fig 7). Les AS sont représentés par des bulles colorées qui sont plus ou moins animés et irrégulières en fonction du nombre d'alertes associées à chaque AS. Des *sparklines* permettent l'inspection rapide des activités et la détection de changements de comportement. En bas de la visualisation, un histogramme montre l'activité globale sur le temps et permet un filtrage temporel. Deux autres vues, une carte globale et une matrice d'activité permettent à la fois une inspection géographique de l'activité et de ses changements.

Comparés aux outils de surveillance, les outils d'analyse offrent une meilleure flexibilité, souvent sans point de départ ou configuration prédéfinis. Lors de la conception de ces outils, une réflexion profonde est nécessaire pour définir les possibilités de cycles d'exploration de données. L'ajout de transitions et d'animations aide à désigner les zones importantes et les modifications de données.

3 La visualisation pour le rapport

Dans certains cas, montrer le résultat visuel lui même peut suffire. Par exemple, quand une alerte est déclenchée par VisAlert, une capture de l'état de la vue pourra rapporter toute l'information nécessaire pour comprendre l'alerte. Quand des anomalies et des alertes sont trouvées pendant la surveillance, ou quand des scénarios d'attaques sont découverts en utilisant des outils d'analyse visuelle, il peut être difficile de communiquer l'idée dans sa globalité. Pour comprendre un scénario complet ou pour expliquer le processus entier d'exploration qui a mené un opérateur à trouver un motif, de nouveaux outils sont nécessaires qui aident à progressivement prendre des notes et construire un rapport compréhensible.

Les auteurs de FlowTag [LC06] partent du postulat que les analyses longues et compliquées de données réseau produisent souvent de mauvais rapports. Leur outil fournit du filtrage simple et de la corrélation par coordonnées parallèles mais surtout la capacité d'étiqueter des éléments intéressants et reliés (Fig 8). Cela facilite la gestion du processus d'analyse mais aussi le partage avec les autres et transforme ce processus en une tâche collaborative. Des données étiquetées facilitent aussi la production de rapports, avec la possibilité de regrouper les données liées.

Pour représenter l'organisation générale de réseaux et le charactère atteignable des différents éléments les composant, Williams et al. utilisent un *tree map* pour représenter les graphes d'attaques combinées [WLI07]. Avec cette technique, des chemins et des scénarios d'attaque potentiels peuvent être représentés de façon plus descriptive.

Alors qu'il pourrait être utilisé en tant qu'outil de surveillance pour l'activité malveillante et les alertes à échelle globale, EMBER [YRB10] est un outil de visualisation dont un des objectifs est la détection d'activités malveillantes et la visualisation d'alertes à grande échelle. Grâce à une configuration fixée de

visualisations classiques, il permet également de communiquer des fréquences d'attaque.

Pour mieux comprendre des attaques complexes comportant de multiples étapes, des visualisations complexes voire des combinaisons de celles-ci sont nécessaires. Yelizarov et al. présentent un outil [YG09] pour visualiser et détecter ces attaques. Positionné dans l'espace, des cylindres représentent des événements, colorés par type et dont la taille correspond à la sévérité. Ces cylindres sont connectés pour montrer les événements liés et séquentiels. Les événements suivent le même axe de temps horizontal sur plusieurs rangées. Ces cylindres sont ensuite reliés à un autre graphe, soit en coordonnées parallèles, soit un nuage de points, qui montre l'adresse source de l'événement.

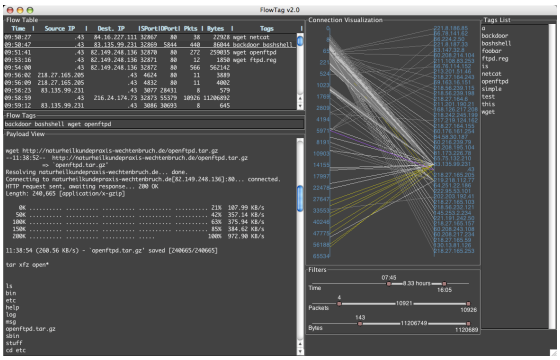

Figure 8. Flowtag [LC06] ajoute la possibilité de faire des annotations pendant ses analyses pour faciliter la production de rapports et leur partage.

En ajoutant aux outils de surveillance et d'analyse des possibilités de prise de note ou de collecte d'une grande quantité de données, les outils pour le rapport facilitent la communication des événements de sécurité qui ont eu lieu.

4 Discussion et conclusion

À notre connaissance, trois autres études ont proposé des taxonomies des outils de visualisation pour la sécurité [Kas06,SSG12,FDCN06]. La première étude [Kas06] classe ces outils suivant le type de données considérées en source (données brutes ou données de plus haut niveau provenant d'IDS, séparées en IDS par signature et IDS comportementaux) et suivant la caractéristique abstraite ou concrète de la visualisation choisie. Les différentes tâches que doivent remplir les outils de visualisation sont évoquées : détecter les activités malveillantes, déterminer les faux positifs, entraîner les IDS par apprentissage. Les deux premières tâches rentrent dans notre catégorie d'outils dédiés à la supervision. La troisième est externe à notre classification. La seconde étude [SSG12] est centrée sur la sécurité réseau et propose une classification par cas d'utilisation : surveillance d'hôtes ou de serveurs, surveillance des connexions entre le réseau interne et le réseau

externe, surveillance de l'activité sur les ports TCP, détection de motifs d'attaques, surveillance des comportements de routage. Les auteurs analysent pour une trentaine d'outils les différentes techniques de visualisation mises en œuvre et les sources de données considérées. Enfin, la troisième étude [FDCN06] propose une classification en deux dimensions des outils de visualisation pour la sécurité réseau. Les différents outils sont classés en fonction de la complexité de la visualisation (du mode texte aux représentations complexes) et de la taille du système représenté (d'une machine à un grand réseau).

Aucune de ces trois autres classifications ne met réellement en avant les objectifs des outils de communications. En ce sens, notre classification se distingue des autres, puisqu'elle repose précisément sur les trois grands objectifs d'un outil de visualisation : la supervision, la fouille visuelle et l'aide à la communication.

Les outils de visualisation en sécurité sont conçus soit pour surveiller des systèmes, soit pour faire des fouilles exploratoires de données de sécurité, soit pour rapporter des résultats. Pour atteindre leurs objectifs, ces outils utilisent différentes approches, en implémentant des combinations caractéristiques de visualisations et d'expériences utilisateur. La conception d'un outil de visualisation va ainsi dépendre à la fois des données d'entrées et de ses objectifs de plus haut niveau. Une majorité d'outils de visualisation pour la sécurité se sont concentrés sur la problématique de la surveillance de systèmes alors que peu d'entre eux se sont intéréssés à l'analyse ou l'aide à la communication. Nous pensons qu'il est important pour les concepteurs d'outils de visualisation de porter leurs efforts sur ces deux objectifs, notamment face à la quantité de données émises par les systèmes d'information de nos jours et à la difficulté de communiquer sur les évènements de sécurité.

Références

AC06. K Abdullah and J A Copeland. Tool Update : High Alarm Count Issues in IDS RainStorm. In *Proc. of VizSEC'06*, pages 129–136, 2006.

AL05. K Abdullah and C Lee. IDS RainStorm : Visualizing IDS Alarms. In *Proc. of VizSEC'05*, pages 1–10, 2005.

Ana. Assisted Analysis. Analyst ' s Notebook 8 Increase the depth of intelligence for effective resource utilization .

CS04. J B Colombe and G Stephens. Statistical Profiling and Visualization for Detection of Malicious Insider Attacks on Computer Networks. In *Proc. of VizSEC/DMSEC'04*, pages 138–142, 2004.

DBWW10. M Daniel, S Bohn, A Wynne, and A William. Real-Time Visualization of Network Behaviors for Situational Awareness. In *Proc. of VizSEC'10*, pages 79–90, 2010.

EW12. S Engle and S Whalen. Visualizing distributed memory computations with hive plots. In *Proc. of VizSEC '12*, pages 56–63, 2012.

FA07. S Foresti and J Agutter. VisAlert : From Idea to Product. In *Proc. of VizSEC'07*, pages 159–174, 2007.

FDCN06. Glenn A Fink, Vyas Duggirala, Ricardo Correa, and Chris North. Bridging the Host-Network Divide : Survey, Taxonomy, and Solution. In *Proc. of LISA '06 : 20th Large Installation System Administration Conference*, pages 247–262, 2006.

FMN05. G A Fink, P Muessig, and C North. Visual Correlation of Host Processes and Network Traffic. In *Proc. of VizSEC'05*, pages 11–19, 2005.

GBT+09. J Glanfield, S Brooks, T Taylor, D Paterson, C Smith, C Gates, and J Mchugh. OverFlow : An Overview Visualization for Network Analysis. In *Proc. of VizSEC'09*, pages 11–19, 2009.

GL05. J R Goodall and W G Lutters. Preserving the Big Picture : Visual Network Traffic Analysis with TNV. In *Proc. of VizSEC'05*, pages 47–54, 2005.

Her06. P Hertzog. Visualizations to Improve Reactivity Towards Security Incidents Inside Corporate Networks. In *Proc. of VizSEC'06*, pages 95–101, 2006.

hiv13. Hive Plots - Linear Layout for Network Visualization - Visually Interpreting Network Structure and Content Made Possible. Technical report, April 2013.

HLW10. L Harrison, A Lu, and W Wang. Interactive Detection of Network Anomalies via Coordinated Multiple Views. In *Proeedings of VizSEC'10*, pages 91–101, 2010.

IVR. B Irwin and J Van Riel. Using InetVis to Evaluate Snort and Bro Scan.

Kas06. R R Kasemri. *A Survey, Taxonomy, and Analysis of Network Security Visualization Techniques.* PhD thesis, 2006.

KFM11. C Kintzel, J Fuchs, and F Mansmann. Monitoring Large IP Spaces with ClockView. In *Proc. of VizSEC'11*, pages 2 :1–2 :10, 2011.

KO04. H Koike and K Ohno. SnortView : Visualization System of Snort Logs. In *Proc. of VizSEC/DMSEC'04*, pages 143–147, 2004.

Koi05. H Koike. Visualizing Cyber Attacks using IP Matrix. In *Proc. of VizSEC'05*, pages 91–98, 2005.

LAL04. K Lakkaraju, E S Ave, and A J Lee. NVisionIP : NetFlow Visualizations of System State for Security Situational Awareness. In *Proc. of VizSEC/DMSEC'04*, pages 65–72, 2004.

LAM+05. Y Livnat, J Agutter, S Moon, R F Erbacher, and S Foresti. A Visualization Paradigm for Network Intrusion Detection. In *Proc. of the Information Assurance Workshop (IAW'05)*, pages 92–99, 2005.

LAMF05. Y Livnat, J Agutter, S Moon, and S Foresti. Visual correlation for situational awareness. In *IEEE Symposium on Information Visualization (INFOVIS'05)*, pages 95–102, 2005.

LC05. C P Lee and J A Copeland. Visual Firewall : Real-time Network Security Monitor Workshop on Visualization for Computer Security. In *Proc. of VizSEC'05*, pages 129–136, 2005.

LC06. C P Lee and J A Copeland. FlowTag : A Collaborative Attack-Analysis, Reporting, and Sharing Tool for Security Researchers. In *Proc. of VizSEC'06*, pages 103–108, 2006.

LMM10. W Lian, F Monrose, and J Mchugh. Traffic Classification Using Visual Motifs : An Empirical Evaluation. 2010.

LSYN05. K Lakkaraju, A Slagell, W Yurcik, and S North. Closing-the-Loop in NVisionIP : Integrating Discovery and Search in Security Visualizations. In *Proc. of VizSEC'05*, pages 75–82, 2005.

MKL04. J Mcpherson, P Krystosk, and L Livermore. PortVis : A Tool for Port-Based Detection of Security Events. In *Proc. of VizSEC/DMSEC'04*, pages 73–81, 2004.

MMB05. C Muelder, K L Ma, and T Bartoletti. A Visualization Methodology for Characterization of Network Scans Workshop on Visualization for Computer Security. In *Proc. of VizSEC'05*, pages 29–38, 2005.

PR. J Pearlman and P Rheingans. Visualizing Network Security Events Using Compound Glyphs from a Service-Oriented.

Ran03. S Ranjan. BGP Eye : A New Visualization Tool for Real-time Detection and Analysis of BGP Anomalies. pages 81–90, 2003.

RCDM+11. F Roveta, G Caviglia, L Di Mario, S Zanero, F Maggi, and P Ciuccarelli. BURN : Baring Unknown Rogue Networks. In *Proc. of VizSEC'11*, pages 6 :1–6 :10, 2011.

RER+10. J Rasmussen, K Ehrlich, S Ross, S Kirk, D Gruen, and J Patterson. Nimble Cybersecurity Incident Management through Visualization and Defensible Recommendations. In *Proc. of VizSEC'10*, pages 102–113, 2010.

Sno13. Snort. Technical report, January 2013.

SSG12. H Shiravi, A Shiravi, and A A Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on Visualization and Computer Graphics*, 18(8) :1313–1329, August 2012.

WLI07. L Williams, R Lippmann, and K Ingols. An Interactive Attack Graph Cascade and Reachability Display. In *Proc. of VizSEC'07*, pages 221–236, 2007.

WMM03. C V Wright, F Monrose, and G M Masson. Using Visual Motifs to Classify Encrypted Traffic. pages 41–50, 2003.

YA04. X Yin and E S Ave. VisFlowConnect : NetFlow Visualizations of Link Relationships for Security Situational Awareness Categories and Subject Descriptors. In *Proc. of VizSEC/DMSEC'04*, pages 26–34, 2004.

YFB+. T H Yu, B W Fuller, J H Bannick, L M Rossey, and R K Cunningham. Integrated Environment Management for Information Operations Testbeds. pages 67–83.

YFB+07. Tamara Yu, Benjamin Fuller, John Bannick, Lee Rossey, and Robert Cunningham. Integrated Environment Management for Information Operations Testbeds LARIAT Provides High-Fidelity Network Emulation and User Simulation. pages 1–25, 2007.

YG09. A Yelizarov and D Gamayunov. Visualization of Complex Attacks and State of Attacked Network. In *Proc. of VizSEC'09*, pages 1–9, 2009.

YRB10. T Yu, J Riordan, and S Boyer. EMBER : A Global Perspective on Extreme Malicious Categories and Subject Descriptors. 2010.

Yur06. W Yurcik. Tool Update : NVisionIP Improvements (Difference View, Sparklines, and Shapes). In *Proc. of VizSEC'06*, pages 65–66, 2006.